

# Causal Inference in Rebuilding and Extending the Recondite Bridge between Finite Population Sampling and Experimental Design

Rahul Mukerjee\*, Tirthankar Dasgupta\*\* and Donald B. Rubin\*\*

\*Indian Institute of Management Calcutta

\*\*Department of Statistics, Harvard University

## Abstract

This article considers causal inference for treatment contrasts from a randomized experiment using potential outcomes in a finite population setting. Adopting a Neymanian repeated sampling approach that integrates such causal inference with finite population survey sampling, an inferential framework is developed for *general* mechanisms of assigning experimental units to multiple treatments. This framework extends classical methods by allowing the possibility of randomization restrictions and unequal replications. Novel conditions that are “milder” than strict additivity of treatment effects, yet permit unbiased estimation of the finite population sampling variance of any treatment contrast estimator, are derived. The consequences of departures from such conditions are also studied under the criterion of minimax bias, and a new justification for using the Neymanian conservative sampling variance estimator in experiments is provided. The proposed approach can readily be extended to the case of treatments with a general factorial structure.

**Keywords:** Potential outcomes, assignment probabilities, treatment contrasts, linear unbiased estimator, stratified assignment, split-plot design.

## 1 Introduction

Finite population sampling and experimental design were two of most researched topics in statistics during the first part of the twentieth century, with both being rooted in the fundamental

statistical principle that inferences should be based only on the randomization distribution actually used, a perspective explicitly first described in Fisher (1925). Ideas from finite population survey sampling were used in the development of the theory of design and analysis of randomized experiments. A pioneering effort along these lines was the seminal work of Neyman (1923/1990), in which he introduced the concept of *potential outcomes* and used it when deriving the sampling, or randomization, distribution of the estimated average treatment effect. During the next few decades, several researchers, e.g., Kempthorne (1955), Wilk (1955), Wilk and Kempthorne (1956, 1957) worked along similar lines. As randomization became more complex, leading to development of designs with randomization restrictions (e.g., block designs, Latin square designs, split-plot designs), derivation of sampling distributions of estimators of treatment effects became increasingly difficult. Plagued by the lack of computational resources to check the correctness of analytical results, these derivations were sometimes incorrect — see Sabbaghi and Rubin (2014) on Neyman (1935).

Development of experimental design methodology in the later part of the twentieth century took a path often dominated by its application to manufacturing and the connection to finite population survey sampling became recondite at best. Because in a controlled manufacturing environment, superpopulation inference is typically of more interest than its finite population counterpart, and a finite population of experimental units can often be viewed as a random sample from a larger population (defined by controllable manufacturing conditions), model-based inference with hypothetical parameters (e.g., the normal linear model) gradually gained popularity and dominated textbooks such as Wu and Hamada (2009).

The potential outcomes framework made a resurgence through its application to causal inference from randomized experiments and observational studies in the social, behavioral, and biomedical (SBB) sciences. Conception and development of this perspective of causal inference can be attributed to the formalization of Neyman’s (Neyman, 1923/1990) notation and results by Rubin (1974, 1975, 1977) in what has been called the “Rubin Causal Model” or RCM (Holland, 1986). However, commonly SBB experiments (see Imbens and Rubin (2015)) involve only a single factor at two levels (e.g., treatment-control studies) and there exists limited literature on the use of the potential outcomes framework to draw inference from experiments with multiple treatments. Such SBB experiments involving multiple treatments, often with a factorial structure, appear to be gaining popularity recently (see, for example, Chakraborty et al. (2009), Collins et al. (2009)).

Recently, Dasgupta et al. (2015) considered the problem of drawing causal inference from balanced  $2^K$  factorial experiments using the potential outcomes framework. Specifically, for a finite population of experimental units, they studied sampling distributions of estimators of factorial effects under a balanced completely randomized treatment assignment mechanism and investigated how the sampling variances of these estimators were affected by departures from the common “Neymanian assumption” of strict additivity of treatment effects (which means all units have exactly the same treatment effect). These results were extended to the case of unbalanced  $2^K$  experiments by Lu (2016). However, completely randomized SBB experiments with exactly two levels for each factor often cannot be conducted. To facilitate application to SBB experiments, it is important to consider assignment mechanisms and treatment structures that are more general than completely randomized  $2^K$  experiments.

This article attempts to reconstruct the nearly broken bridge between finite population survey sampling and experimental design with the aim of developing a framework for causal inference for treatment contrasts from randomized experiments with a *general assignment mechanism* within a finite population setting. The assignment mechanisms considered allow the possibility of randomization restrictions, such as stratification or split-plot structure, and unequal replications. A new result shows that unbiased estimation of the sampling variance of any treatment contrast estimator is possible under conditions milder than standard (e.g., without strict additivity). The consequences of departures from such conditions are also studied, and a new justification behind using the so-called conservative “Neymanian variance estimator” of the estimated treatment contrast is provided by considering minimax bias. The proposed approach applies readily to the situation where the treatments have a *general* factorial structure with  $K$  factors, each with a possibly different number of levels.

The article is organized as follows: Section 2 gives a brief overview of potential outcomes and associated estimands (quantities of interest) at the unit-level and the population-level. In Section 3, a general assignment mechanism is introduced and illustrated using three examples that explain its connection to finite population sampling theory. Section 4 develops the Neymanian inference procedure under the general assignment mechanism and presents results that identify conditions, milder than strict additivity, permitting unbiased estimation of the sampling variances (and covariances) of linear estimators of treatment contrasts. These conditions are examined in Section 5, and two examples are used to illustrate their applicability. Section 6 explores the

effect of departures from additivity (strict and mild) on the bias of the estimators and identifies the Neymanian variance estimator as a minimax solution. Section 7 shows how the proposed framework can be applied to factorial experiments, and Section 8 illustrates the proposed approach using some simulated numerical examples. Section 9 presents a discussion and indicates future research.

## 2 Potential outcomes, unit-level and population-level treatment contrasts

Consider a randomized experiment that involves allocation of a set  $Z$  of treatments to a population  $\{1, \dots, N\}$  of  $N$  experimental units such that each unit is assigned to one of the  $|Z|$  treatments ( $|Z|$  denotes the cardinality of the set  $Z$ ). For each  $z \in Z$  and  $i = 1, \dots, N$ , let  $Y_i(z)$  denote the potential outcome of unit  $i$  if assigned to treatment  $z$ . A unit-level treatment contrast for unit  $i$  is of the form:

$$\tau_i = \sum_{z \in Z} g(z) Y_i(z), \quad (1)$$

where  $g(z)$ ,  $z \in Z$ , are known, not all zeros, and sum to zero. Such a contrast is a unit-level causal estimand under the RCM that compares the potential outcomes of a unit when exposed to different treatments. Let  $\bar{Y}(z) = \sum_{i=1}^N Y_i(z)/N$ ,  $z \in Z$  denote the mean of the  $N$  potential outcomes for treatment  $z$  across all units. Then the mean of the unit-level estimands  $\tau_1, \dots, \tau_N$ , namely,

$$\bar{\tau} = (\tau_1 + \dots + \tau_N)/N = \sum_{z \in Z} g(z) \bar{Y}(z), \quad (2)$$

defines a population-level treatment contrast, which is a causal estimand for the finite population of units. Our interest lies in Neymanian inference (Rubin, 1990) for treatment contrasts, like  $\bar{\tau}$ , on the basis of the outcomes observed from the experiment. If the treatments have a factorial structure, then treatment contrasts representing the factorial main effects and interactions are typically of interest. Such contrasts will be defined in Section 7.

## 3 A general assignment mechanism for randomized experiments

An assignment mechanism is the rule that assigns experimental units to treatments; see Imbens and Rubin (2015) for a detailed discussion of assignment mechanisms. Here we consider a general

assignment mechanism for randomized experiments that is inspired by a general sampling design for a finite population (e.g., Hansen et al. (1953), Kish (1965), Cochran (1977)). Let  $T = \{T(z) : z \in Z\}$  be any partition of the population  $\{1, \dots, N\}$  of experimental units into  $|Z|$  nonempty disjoint subsets  $T(z), z \in Z$ . We consider a general mechanism for assigning units to treatments that involves the following two steps:

- (i) Select a partition of the  $N$  units,  $T = \{T(z) : z \in Z\}$ , where the partition has known probability  $p(T) (\geq 0)$  of being selected, and
- (ii) assign all the units in  $T(z)$  to treatment  $z$ , for every  $z \in Z$ .

In the remaining part of this paper, we shall refer to the above assignment mechanism as a *general assignment mechanism* (GAM), which must satisfy  $\sum_T p(T) = 1$ , where  $\sum_T$  denotes the sum over all such partitions  $T$ . In the three examples discussed in Section 3.1, all  $T$  with  $p(T) > 0$  have the same  $p(T)$ , but this is irrelevant to the theory developed here. Implicitly, all probabilities are conditional in the sense that they treat the values of all potential outcomes and blocking factors as fixed.

Under a GAM, the observed outcome for unit  $i$  can be expressed as:

$$Y_i^{\text{obs}} = \sum_{z \in Z} W_i(z) Y_i(z),$$

where  $W_i(z)$  is a binary random variable satisfying:

$$Pr(W_i(z) = 1) = E(W_i(z)) = Pr(i \in T(z)), z \in Z;$$

that is,  $W_i(z)$  takes value 1 if unit  $i$  is assigned to treatment  $z$  and zero otherwise.

### 3.1 Three examples

**Example 1: Stratified random assignment.** Let the population  $\{1, \dots, N\}$  of units be partitioned into  $H$  strata  $\Omega_1, \dots, \Omega_H$  with  $|\Omega_h| = N_h (\geq 2)$  for  $h = 1, \dots, H$ , and  $\sum_{h=1}^H N_h = N$ . Consider a randomization, that assigns  $r_h(z)$  units in  $\Omega_h$  to treatment  $z$ , where  $r_h(z) (\geq 2)$ ,  $z \in Z$ , are fixed integers that sum to  $N_h$  ( $h = 1, \dots, H$ ) and all assignments within each stratum are equiprobable.

Such an assignment mechanism can be equivalently described as follows: for  $h = 1, \dots, H$ , let  $T_h(z)$  denote the set of  $r_h(z)$  units assigned to treatment  $z$  in stratum  $\Omega_h$ . Then  $T(z) =$

$T_1(z) \cup \dots \cup T_H(z)$  is the set of  $\sum_{h=1}^H r_h(z)$  units from all the strata assigned to treatment  $z$ , and  $\{T(z) : z \in Z\}$  is a partition of the  $N$  units into  $|Z|$  treatment groups. The assignment mechanism selects one among all such possible partitions with equal probability. In particular, if there is a single stratum, i.e.,  $H = 1$ , then the assignment is completely randomized, as defined in textbooks in classical experimental design.

**Example 2: Split-plot assignment.** Assume that the treatments have a factorial structure, representing treatment combinations of two factors  $F_1$  and  $F_2$ , each at two or more levels. Let  $Z_k$  denote the set of levels of  $F_k$  ( $k = 1, 2$ ), and let  $z = z_1 z_2$  denote a treatment combination, where  $z_k \in Z_k$ . In many practical situations, the levels of  $F_1$  are more difficult to change from unit to unit than the levels of  $F_2$ . In such situations, it can be advantageous to employ a “split-plot” assignment of units to treatment combinations. Let  $N = H N_0$ , for two integers  $H$  and  $N_0$ , both larger than one, and partition the population of  $N$  units into  $H$  disjoint sets  $\Omega_1, \dots, \Omega_H$ , which are called whole-plots, each of which consists of  $N_0$  units, called sub-plots nested within each whole-plot. Consider now a two-stage randomization of treatments, that assigns  $r_1(z_1)$  ( $\geq 2$ ) whole-plots to level  $z_1$  of  $F_1$ , and then, within each whole-plot, assigns  $r_2(z_2)$  sub-plots to level  $z_2$  of  $F_2$ . Here, we suppose that, first, all assignments at each stage are equiprobable; second, that the fixed positive integers,  $r_1(z_1)$ ,  $z_1 \in Z_1$ , sum to  $H$ ; and third, that the fixed positive integers  $r_2(z_2)$ ,  $z_2 \in Z_2$ , sum to  $N_0$ . This is equivalent to selecting, with equal probability, one of the partitions of the form  $\{T(z_1 z_2) : z_1 z_2 \in Z\}$ , where

$$T(z_1 z_2) = \bigcup_{h \in T_1(z_1)} T_{h2}(z_2), \quad (3)$$

for each  $z_1 z_2$ . Here  $T_1(z_1)$  represents the set of whole-plots assigned to level  $z_1$  of factor  $F_1$ , so that for  $z_1 \in Z_1$ , the sets  $T_1(z_1)$  are disjoint having union  $\{1, \dots, H\}$ . The set  $T_{h2}(z_2)$  represents the set of sub-plots within whole-plot  $\Omega_h$  that are assigned to level  $z_2$  of factor  $F_2$ . For each  $h = 1, \dots, H$ , the sets  $T_{h2}(z_2)$ ,  $z_2 \in Z_2$ , are also disjoint having union  $\Omega_h$ . Moreover,  $|T_1(z_1)| = r_1(z_1)$  and  $|T_{h2}(z_2)| = r_2(z_2)$ . It is important to note that, although the whole-plots  $\Omega_1, \dots, \Omega_H$  are formally similar to the strata in Example 1, the assignment mechanism is now different because of the two-stage randomization involved.

**Example 3: Unicluster assignment.** Partition the population of  $N$  units into  $|Z|$  disjoint clusters  $\Delta(1), \dots, \Delta(|Z|)$ . Analogous to unicluster designs in finite population sampling (Hedayat

and Sinha, 1991) which include systematic sampling, consider an assignment mechanism that assigns all units in a cluster to one of the treatments completely at random, that is, the treatments assigned to clusters  $\Delta(1), \dots, \Delta(|Z|)$  form a random permutation of  $1, \dots, |Z|$ , all such permutations being equiprobable. This is equivalent to selecting, with equal probability, one of the  $|Z|!$  partitions  $\{T(z) : z \in Z\}$ , where each  $T(z)$  equals one of the clusters.

### 3.2 Assignment probabilities of first two orders

Recall that  $W_i(z)$  denotes the binary treatment indicator that unit  $i$  receives treatment  $z$ . The assignment probabilities of the first two orders are then given by  $\pi_i(z) = E[W_i(z)]$  and  $\pi_{ii^*}(z, z^*) = E[W_i(z)W_{i^*}(z^*)]$ , where  $i, i^* (\neq i) = 1, \dots, N$ , and  $z, z^* \in Z$ . Thus,  $\pi_i(z)$  is the probability that unit  $i$  is assigned to treatment  $z$ . Similarly,  $\pi_{ii^*}(z, z^*)$  is the probability that unit  $i$  is assigned to treatment  $z$  and unit  $i^*$  is assigned to treatment  $z^*$ . Much of the theoretical developments under a GAM that will be developed in the next few sections depends on these quantities.

Recall that our interest lies in estimation of linear combinations (contrasts) of treatment means  $\bar{Y}(z), z \in Z$ . In finite population sampling (e.g., see Kish, 1965) a necessary and sufficient condition to allow unbiased estimation of these treatment means is:

$$\pi_i(z) > 0, \quad i = 1, \dots, N, \quad z \in Z, \quad (4)$$

where these  $\pi_i(z)$  are known. Thus the GAM will, hereafter, be assumed to satisfy condition (4). Conditions involving  $\pi_{ii^*}(z, z^*)$  will be imposed in Section 4 for estimation of sampling variance.

The assignment probabilities for the assignment mechanisms in the three examples introduced in Section 3.1 are shown in the Appendix.

## 4 Neymanian inference for treatment contrasts

We now focus on our main objective, i.e., inferring the values of finite population treatment contrasts defined in (2), i.e., Neymanian randomization-based inference for the finite population treatment contrasts, based on evaluation of expectations of statistics over the distribution induced by a GAM (Rubin, 2008). From (2), it is clear that an unbiased estimator of the finite population contrast  $\bar{\tau}$  is the obvious method of moments estimator

$$\hat{\bar{\tau}} = \sum_{z \in Z} g(z) \hat{Y}(z), \quad (5)$$

where  $\widehat{Y}(z)$  is an unbiased estimator of  $\bar{Y}(z)$ . In Section 4 we will compute the sampling properties of this estimator under a GAM. Specifically, in Section 4.1, we will introduce linear unbiased estimators (LUE) of  $\bar{Y}(z)$  from the finite population sampling literature, substitute the LUE in (5) to obtain an LUE of  $\bar{\tau}$ , and obtain an analytical expression for its sampling variance with respect to the distribution induced by the GAM. In the subsequent Sections 4.2 and 4.3, we discuss the estimation of sampling variance and covariance, respectively, and introduce a new approach.

#### 4.1 LUEs of treatment contrasts and their sampling variances

For a GAM, let

$$\widehat{Y}(z) = a(T, z) + \sum_{i \in T(z)} b_i(T, z) Y_i(z), \quad (6)$$

be any LUE of the treatment mean  $\bar{Y}(z)$ ,  $z \in Z$ , where  $a(T, z)$  and  $b_i(T, z)$  are known quantities such that  $a(T, z)$  depends on  $z$  and the selected partition  $T$ , and  $b_i(T, z)$  depends, also, on  $i$ . An example of such a LUE of  $\bar{Y}(z)$  is the popular Horvitz-Thompson estimator (Horvitz and Thompson, 1952), which is a kind of weighted average of the observed outcomes of all units exposed to treatment  $z$ , where the weight for each unit is the inverse of its first-order assignment probability, i.e., in (6),  $a(T, z) = 0$  and  $b_i(T, z) = (N\pi_i(z))^{-1}$ . This estimator, henceforth denoted by  $\widehat{Y}^{\text{HT}}(z)$ , is

$$\widehat{Y}^{\text{HT}}(z) = \sum_{i \in T(z)} \frac{Y_i(z)}{N\pi_i(z)}. \quad (7)$$

Given a LUE  $\widehat{Y}(z)$  of the form (6), a LUE  $\widehat{\tau}$  of the estimand of interest,  $\bar{\tau}$ , can be obtained by substituting  $\widehat{Y}(z)$  from (6) into (5). The sampling variance of  $\widehat{\tau}$  is

$$\text{var}(\widehat{\tau}) = E(\widehat{\tau}^2) - \bar{\tau}^2 = \sum_{z \in Z} \sum_{z^* \in Z} g(z)g(z^*) E\left(\widehat{Y}(z)\widehat{Y}(z^*)\right) - \bar{\tau}^2. \quad (8)$$

To calculate the first term on the right hand side of (8), we use the following lemma, that is a standard result in finite population sampling literature (Chaudhuri, 2014):

**Lemma 1.** *If  $\widehat{Y}(z)$  is a LUE of  $\bar{Y}(z)$  and is of the form (6), then*

$$\begin{aligned} E\left(\widehat{Y}(z)\widehat{Y}(z^*)\right) &= A(z, z^*) + \sum_{i=1}^N \left(A_i^{(1)}(z, z^*) Y_i(z) + A_i^{(2)}(z, z^*) Y_i(z^*)\right) \\ &+ \sum_{i=1}^N \sum_{i^*=1}^N B_{ii^*}(z, z^*) Y_i(z) Y_{i^*}(z^*), \end{aligned} \quad (9)$$



where  $A(z, z^*)$ ,  $A_i^{(1)}(z, z^*)$ ,  $A_i^{(2)}(z, z^*)$  and  $B_{ii^*}(z, z^*)$  are known constants that may depend on  $z$  and  $z^*$  as well as the respective subscripts. In particular,

$$B_{ii^*}(z, z^*) = \sum_{T: i \in T(z), i^* \in T(z^*)} p(T) b_i(T, z) b_{i^*}(T, z^*). \quad (10)$$

In a GAM, because unit  $i$  cannot be assigned to both  $z$  and  $z^*$ , unit  $i$  cannot be in  $T(z)$  and  $T(z^*)$ . Therefore for  $i = 1, \dots, N$ , the term  $B_{ii^*}(z, z^*)$  in Lemma 1 vanishes whenever  $z \neq z^*$ . The following result can now be established using (8) and equation (9) of Lemma 1.

**Theorem 1.** *Under a GAM, the sampling variance of  $\hat{\tau}$  is:*

$$\begin{aligned} \text{var}(\hat{\tau}) &= M + \sum_{z \in Z} \sum_{i=1}^N \left( M_i(z) Y_i(z) + M_{ii}(z) (Y_i(z))^2 \right) \\ &\quad + \sum_{z \in Z} \sum_{z^* \in Z} \sum_{i=1}^N \sum_{i^* (\neq i)=1}^N M_{ii^*}(z, z^*) Y_i(z) Y_{i^*}(z^*) - \bar{\tau}^2, \end{aligned} \quad (11)$$

where

$$\begin{aligned} M &= \sum_{z \in Z} \sum_{z^* \in Z} g(z) g(z^*) A(z, z^*), & M_i(z) &= g(z) \sum_{z^* \in Z} g(z^*) \left( A_i^{(1)}(z, z^*) + A_i^{(2)}(z^*, z) \right), \\ M_{ii}(z) &= (g(z))^2 B_{ii}(z, z), & M_{ii^*}(z, z^*) &= g(z) g(z^*) B_{ii^*}(z, z^*), \end{aligned} \quad (12)$$

for  $z, z^* \in Z$  and  $i, i^* (\neq i) = 1, \dots, N$ .

*Remark 1.* Theorem 1, and the later results, generalize several results on the sampling variance of estimators of causal estimands available in the experimental design and causal inference literature. Consider the important special case of our formulation with two treatments,  $Z = \{0, 1\}$ . Let the estimand of interest be the average treatment effect  $\bar{\tau} = \bar{Y}(1) - \bar{Y}(0)$ , so that  $g(1) = 1$  and  $g(0) = -1$ . Consider the stratified assignment mechanism described in Example 1 of Section 3.1 with  $H = 1$ , i.e., a completely randomized assignment. Let  $r(1) (\geq 2)$  and  $r(0) (\geq 2)$  units be exposed to treatments 1 and 0 respectively, and consider the HT-estimator  $\hat{\tau}^{\text{HT}} = \sum_{i \in T(1)} Y_i(1)/r(1) - \sum_{i \in T(0)} Y_i(0)/r(0)$  as a LUE of  $\bar{\tau}$ . Then, after some non-trivial algebraic simplifications that are similar to results discussed later in Section 5.4, the following expression for the sampling variance of  $\hat{\tau}^{\text{HT}}$  is obtained from (11) :

$$\begin{aligned} \text{var}(\hat{\tau}^{\text{HT}}) &= \frac{\sum_{i=1}^N (Y_i(0) - \bar{Y}(0))^2}{(N-1)r(0)} + \frac{\sum_{i=1}^N (Y_i(1) - \bar{Y}(1))^2}{(N-1)r(1)} - \frac{\sum_{i=1}^N (\tau_i - \bar{\tau})^2}{N(N-1)} \\ &= \frac{S(0,0)}{r(0)} + \frac{S(1,1)}{r(1)} - \frac{S(\tau, \tau)}{N}, \end{aligned} \quad (13)$$

where  $S(z, z)$  is the variance of  $Y_1(z), \dots, Y_N(z)$  with divisor  $N - 1$ , and  $S(\tau, \tau)$  is the variance of unit-level treatment effects  $\tau_1, \dots, \tau_N$  also with divisor  $N - 1$ .

The decomposition of this sampling variance into two components:  $S(0, 0)/r(0) + S(1, 1)/r(1)$  and  $S(\tau, \tau)/N$  in (13) above, was originally developed by Neyman (Neyman, 1923/1990), and is a much studied result in causal inference literature (see, for example Rubin (1980), Rubin (1990), Gadbury (2001), Ch. 7 of Imbens and Rubin (2015), Dasgupta et al. (2015)). This decomposition plays a crucial role in understanding the conditions under which  $\text{var}(\hat{\tau}^{\text{HT}})$  can be unbiasedly estimated. Henceforth we refer to this decomposition as the ‘‘Neymanian decomposition’’.

## 4.2 Estimation of sampling variance

Even if (4) holds and, in addition, the second-order assignment probabilities are all positive,  $\text{var}(\hat{\tau})$  does not admit an unbiased estimator because, by (2), the last term,  $\bar{\tau}^2$ , in (11) involves products like  $Y_i(z)Y_i(z^*)$ , and one can never observe both  $Y_i(z)$  and  $Y_i(z^*)$  for  $z \neq z^*$ . This problem with estimation of the sampling variance was first noted by Neyman (Neyman, 1923/1990), which suggested that unbiased estimation of the sampling variance is only generally possible under the assumption of *strict additivity*. Additivity and its impact on estimation of sampling variance of  $\hat{\tau}$  has been studied by several authors since 1923. Recently, Dasgupta et al. (2015) showed that for a completely randomized balanced  $2^K$  factorial experiment, the Neymanian estimator of  $\hat{\tau}$  has an upward bias that vanishes under strict additivity.

**Definition 1.** The potential outcomes  $Y_i(z), i = 1, \dots, N, z \in Z$ , have *strictly additive* treatment effects if for every  $z, z^* \in Z$ ,  $Y_i(z) - Y_i(z^*)$  is the same for  $i = 1, \dots, N$ , or equivalently, in matrix notation,

$$(\mathbf{I} - N^{-1}\mathbf{J})[\mathbf{Y}(z) - \mathbf{Y}(z^*)] = \mathbf{0}, \quad z, z^* \in Z, \quad (14)$$

where  $\mathbf{I}$  is the identity matrix and  $\mathbf{J}$  is the matrix of all ones, both of order  $N$ , and  $\mathbf{Y}(z)$  is the column vector with elements  $Y_1(z), \dots, Y_N(z)$  for each  $z \in Z$ .

In our general formulation, which covers the one in Dasgupta et al. (2015), we explore unbiased estimation of  $\text{var}(\hat{\tau})$  under conditions that are milder than strict additivity. In what follows, any positive semidefinite (psd) matrix is assumed, as usual, to be also symmetric.

Let  $\mathbb{Q}$  denote a class of psd matrices  $\mathbf{Q}$  of order  $N$  with  $(i, i^*)$ th element  $q_{ii^*}$ , such that the

$q_{ii^*}$  are known constants and

$$\mathbf{Q}\mathbf{J} = \mathbf{0}, \quad q_{ii} = 1/N^2, \quad i = 1, \dots, N. \quad (15)$$

An example is:

$$\mathbf{Q}_{\text{strict}} = (N(N-1))^{-1} (\mathbf{I} - N^{-1}\mathbf{J}). \quad (16)$$

*Remark 2.* The connection between  $\mathbf{Q}_{\text{strict}}$  and the Neymanian decomposition arises because the last term in (13) can be written as:

$$\frac{S(\tau, \tau)}{N} = \frac{1}{N(N-1)} \sum_{i=1}^N (\tau_i - \bar{\tau})^2 = \boldsymbol{\tau}' \mathbf{Q}_{\text{strict}} \boldsymbol{\tau},$$

where  $\boldsymbol{\tau} = (\tau_1, \dots, \tau_N)'$  is the vector of unit-level treatment contrasts. Also, from Definition 1, the condition for strict additivity can be stated equivalently as

$$\mathbf{Q}_{\text{strict}}[\mathbf{Y}(z) - \mathbf{Y}(z^*)] = \mathbf{0}, \quad z, z^* \in Z.$$

We now show how the class of matrices  $\mathbb{Q}$  plays a crucial role in “extracting” an estimable component of  $\text{var}(\hat{\tau})$  and thus leads to a generalization of the Neymanian decomposition. Consider the quantity  $M_{ii^*}(z, z^*)$  defined in the last line of (12) that appears in the third term of (11), and modify it:

$$\widetilde{M}_{ii^*}(z, z^*) = g(z)g(z^*) \{B_{ii^*}(z, z^*) + q_{ii^*} - (1/N^2)\}, \quad (17)$$

where  $q_{ii^*}$  is the  $(i, i^*)$ th element of  $\mathbf{Q} \in \mathbb{Q}$ . We now replace  $M_{ii^*}(z, z^*)$  by  $\widetilde{M}_{ii^*}(z, z^*)$  in (11), and drop the term  $\bar{\tau}^2$  to define, for  $\mathbf{Q} \in \mathbb{Q}$ ,

$$V_{\mathbf{Q}}(\hat{\tau}) = M + \sum_{z \in Z} \sum_{i=1}^N \left( M_i(z) Y_i(z) + M_{ii}(z) (Y_i(z))^2 \right) + \sum_{z \in Z} \sum_{z^* \in Z} \sum_{i=1}^N \sum_{i^* (\neq i)=1}^N \widetilde{M}_{ii^*}(z, z^*) Y_i(z) Y_{i^*}(z^*). \quad (18)$$

Note that the quantity  $V_{\mathbf{Q}}(\hat{\tau})$  depends on  $\mathbf{Q} \in \mathbb{Q}$  through the  $\widetilde{M}_{ii^*}(z, z^*)$ 's, which depend on  $\mathbf{Q}$ .

**Theorem 2.** For any  $\mathbf{Q} \in \mathbb{Q}$ , where  $\mathbb{Q}$  represents a class of psd matrices of order  $N$  satisfying (15),

(1) The sampling variance of  $\hat{\tau}$  given by (11) of Theorem 1 can be decomposed as:

$$\text{var}(\hat{\tau}) = V_{\mathbf{Q}}(\hat{\tau}) - \boldsymbol{\tau}' \mathbf{Q} \boldsymbol{\tau}, \quad (19)$$

where  $\boldsymbol{\tau} = (\tau_1, \dots, \tau_N)'$  is the vector of unit-level treatment contrasts given by (1), and  $V_{\mathbf{Q}}(\hat{\tau})$  is given by (18).

(2) An upper bound of  $\text{var}(\hat{\tau})$  is given by  $V_{\mathbf{Q}}(\hat{\tau})$ .

(3) This bound is attained for every treatment contrast  $\bar{\tau}$  if and only if:

$$\mathbf{Q} \{ \mathbf{Y}(z) - \mathbf{Y}(z^*) \} = 0, \quad z, z^* \in Z. \quad (20)$$

The proof of Theorem 2 is in the Appendix.

*Remark 3.* Theorem 2 has the following implications:

- (i) Part (1) of Theorem 2 provides a generalization of the Neymanian decomposition of  $\text{var}(\hat{\tau})$  through the class of matrices  $\mathbb{Q}$ . Because the Neymanian decomposition is achieved using the matrix  $\mathbf{Q}_{\text{strict}}$  defined by (16), and  $\mathbf{Q}_{\text{strict}} \in \mathbb{Q}$  as noted earlier, it follows that the Neymanian decomposition is a special case of the decomposition (19) in part (1).
- (ii) By standard arguments in finite population sampling, the component  $V_{\mathbf{Q}}(\hat{\tau})$  in the decomposition of  $\text{var}(\hat{\tau})$ , given by (18), can be unbiasedly estimated by the quantity:

$$\begin{aligned} \hat{V}_{\mathbf{Q}}(\hat{\tau}) &= M + \sum_{z \in Z} \sum_{i \in T(z)} \frac{1}{\pi_i(z)} \left( M_i(z) Y_i(z) + M_{ii}(z) (Y_i(z))^2 \right) \\ &\quad + \sum_{z \in Z} \sum_{z^* \in Z} \sum_{i \in T(z)} \sum_{i^* \in T(z^*)} \frac{\widetilde{M}_{ii^*}(z, z^*)}{\pi_{ii^*}(z, z^*)} Y_i(z) Y_{i^*}(z^*), \end{aligned} \quad (21)$$

provided  $\pi_i(z) > 0$ , which is satisfied by assumption (4), and

$$\pi_{ii^*}(z, z^*) > 0, \quad \text{whenever } \widetilde{M}_{ii^*}(z, z^*) \neq 0. \quad (22)$$

Part (2) of Theorem 2 thus implies that  $\hat{V}_{\mathbf{Q}}(\hat{\tau})$  overestimates  $\text{var}(\hat{\tau})$ .

- (iii) Part (3) of Theorem 2 implies that for any  $\mathbf{Q} \in \mathbb{Q}$  satisfying (20),  $\text{var}(\hat{\tau})$  equals  $V_{\mathbf{Q}}(\hat{\tau})$ , and thus can be unbiasedly estimated by  $\hat{V}_{\mathbf{Q}}(\hat{\tau})$  given by (21), provided (22) holds. Remark 2 implies that condition (20), with  $\mathbf{Q} = \mathbf{Q}_{\text{strict}}$ , is equivalent to strict additivity. For any other  $\mathbf{Q} \in \mathbb{Q}$ , it is, therefore, immediate that condition (20) is no more demanding than strict additivity because, by (16) and the first equation in (15), the rows of  $\mathbf{Q}$  are spanned by those of  $\mathbf{Q}_{\text{strict}}$ . Indeed, for any such  $\mathbf{Q}$ , (20) is a milder condition than strict additivity whenever the row space of  $\mathbf{Q}$  is a proper subspace of the row space of  $\mathbf{Q}_{\text{strict}}$ , i.e.,  $\text{rank}(\mathbf{Q}) < N - 1$ . Thus, (20) of Theorem 2 provides a condition that is milder than strict additivity but still permits unbiased estimation of  $\text{var}(\hat{\tau})$ .

The points noted in Remark 3 underscore the central role played by (20) and (22) in unbiased sampling variance estimation. Condition (22) leads to the estimator (21), which is unbiased for  $V_{\mathbf{Q}}(\hat{\tau})$  but has a non-negative bias for  $\text{var}(\hat{\tau})$ . Under condition (20), (21) is unbiased for  $\text{var}(\hat{\tau})$  as well. A crucial difference between these conditions is that condition (20) involves the potential outcomes but not the assignment mechanism, and condition (22) involves the assignment mechanism but not the potential outcomes. Both conditions, however, involve the matrix  $\mathbf{Q} \in \mathbb{Q}$ . Henceforth, we will refer to condition (22) as the ***second-order assignment probability condition*** (SAP condition) and condition (20) as the ***generalized additivity condition*** (GA condition). We examine both of these conditions in more detail in the context of Examples 1-3 in Section 5.

### 4.3 Sampling covariance and its estimation

The ideas in Section 4.2 also apply to the estimation of sampling covariance. Consider two treatment contrasts  $\bar{\tau}_l = \sum_{z \in Z} g_l(z) \bar{Y}(z)$  and their LUEs  $\hat{\tau}_l$ ,  $l = 1, 2$ . Using Lemma 1, we have Theorem 3, which parallels Theorem 1:

**Theorem 3.** *Under a GAM, the sampling covariance between  $\hat{\tau}_1$  and  $\hat{\tau}_2$  is:*

$$\begin{aligned} \text{cov}(\hat{\tau}_1, \hat{\tau}_2) &= R + \sum_{z \in Z} \sum_{i=1}^N \left( R_i(z) Y_i(z) + R_{ii}(z) (Y_i(z))^2 \right) \\ &\quad + \sum_{z \in Z} \sum_{z^* \in Z} \sum_{i=1}^N \sum_{i^* (\neq i)=1}^N R_{ii^*}(z, z^*) Y_i(z) Y_{i^*}(z^*) - \bar{\tau}_1 \bar{\tau}_2, \end{aligned} \quad (23)$$

where

$$\begin{aligned} R &= \sum_{z \in Z} \sum_{z^* \in Z} g_1(z) g_2(z^*) A(z, z^*), \\ R_i(z) &= g_1(z) \sum_{z^* \in Z} g_2(z^*) A_i^{(1)}(z, z^*) + g_2(z) \sum_{z^* \in Z} g_1(z^*) A_i^{(2)}(z^*, z), \\ R_{ii}(z) &= g_1(z) g_2(z) B_{ii}(z, z), \quad R_{ii^*}(z, z^*) = g_1(z) g_2(z^*) B_{ii^*}(z, z^*), \end{aligned} \quad (24)$$

for  $z, z^* \in Z$  and  $i, i^* (\neq i) = 1, \dots, N$ .

An unbiased estimator of  $\text{cov}(\hat{\tau}_1, \hat{\tau}_2)$  can be derived using ideas similar to those used in Section 4.2 for unbiased estimation of the sampling variance of  $\hat{\tau}$ . For any  $\mathbf{Q} \in \mathbb{Q}$ , define, similar to (17),

$$\tilde{R}_{ii^*}(z, z^*) = g_1(z) g_2(z^*) \{ B_{ii^*}(z, z^*) + q_{ii^*} - (1/N^2) \}. \quad (25)$$

Proceed, as in the proof of Theorem 2, part (1), to obtain

$$\text{cov}(\hat{\tau}_1, \hat{\tau}_2) = C_{\mathbf{Q}}(\hat{\tau}_1, \hat{\tau}_2) - \boldsymbol{\tau}'_{[1]} \mathbf{Q} \boldsymbol{\tau}_{[2]},$$

where  $\boldsymbol{\tau}_{[1]}$  and  $\boldsymbol{\tau}_{[2]}$  are column vectors of unit-level treatment contrasts corresponding to  $\bar{\tau}_1$  and  $\bar{\tau}_2$  respectively, and

$$\begin{aligned} C_{\mathbf{Q}}(\hat{\tau}_1, \hat{\tau}_2) = & \mathbf{R} + \sum_{z \in Z} \sum_{i=1}^N \left( R_i(z) Y_i(z) + R_{ii}(z) (Y_i(z))^2 \right) \\ & + \sum_{z \in Z} \sum_{z^* \in Z} \sum_{i=1}^N \sum_{i^* (\neq i)=1}^N \tilde{R}_{ii^*}(z, z^*) Y_i(z) Y_{i^*}(z^*). \end{aligned}$$

In the spirit of the SAP condition (22), if  $\pi_{ii^*}(z, z^*) > 0$ , whenever  $\tilde{R}_{ii^*}(z, z^*) \neq 0$ , then  $C_{\mathbf{Q}}(\hat{\tau}_1, \hat{\tau}_2)$  admits an unbiased estimator along the lines of (21). If, in addition, the GA condition (20) holds, then  $\boldsymbol{\tau}'_{[1]} \mathbf{Q} \boldsymbol{\tau}_{[2]}$  vanishes and this estimator becomes unbiased for  $\text{cov}(\hat{\tau}_1, \hat{\tau}_2)$  as well. Thus GA plays a crucial role in unbiased estimation of, not only the sampling variance of each treatment contrast estimator, but also the sampling covariance between two of them

## 5 A closer examination of the unbiasedness conditions, with examples

In Section 4, we identified two key conditions that allow unbiased estimation of the sampling variance of  $\hat{\tau}$ . The GA condition is a property of the unknown potential outcomes that cannot be known even after experimentation. In contrast, the SAP condition depends on the assignment mechanism. In this section, we explore these two conditions in more depth with respect to the examples in Section 3.1.

### 5.1 The GA condition and choice of $\mathbf{Q}$

The GA condition, given by (20), is fulfilled if there exists a matrix  $\mathbf{Q} \in \mathbb{Q}$ , such as  $\mathbf{Q}_{\text{strict}}$  in (16), that satisfies  $\mathbf{Q}(\mathbf{Y}(z) - \mathbf{Y}(z^*)) = 0$ ,  $z, z^* \in Z$ . We now explore some other choices of  $\mathbf{Q}$ .

- (a) **Stratum-level additivity:** Let  $N_1, \dots, N_H$  ( $\geq 2$ ) be integers that sum to  $N$ , and for  $h = 1, \dots, H$ , let  $\mathbf{Q}_h$  be the square matrix of order  $N_h$  having each diagonal element equal to  $1/N^2$  and each off-diagonal element equal to  $-1/(N^2(N_h - 1))$ . Then the psd matrix

$$\mathbf{Q}_{\text{strat}} = \text{diag}(\mathbf{Q}_1, \dots, \mathbf{Q}_H) \tag{26}$$

satisfies (15) and therefore  $\mathbf{Q}_{\text{strat}} \in \mathbb{Q}$ . Let  $\Omega_1 = \{1, \dots, N_1\}, \Omega_2 = \{N_1 + 1, \dots, N_1 + N_2\}, \dots, \Omega_H = \{\sum_{h=1}^{H-1} N_h + 1, \dots, N\}$ . For  $\mathbf{Q}_{\text{strat}}$  defined in (26), condition (20) is equivalent to having additivity within each  $\Omega_h$  which, for  $H \geq 2$ , is milder than strict additivity.

- (b) **Whole-plot-level additivity:** Let  $H \geq 2$ . With  $N_1, \dots, N_H$  and  $\Omega_1, \dots, \Omega_H$  as in (a), now suppose  $N$  is an integral multiple of  $H$  and let  $N_1 = \dots = N_H = N/H = N_0$ . Take  $\mathbf{Q}$  in partitioned form as:

$$\mathbf{Q}_{\text{whole-plot}} = (\mathbf{Q}_{hh^*})_{h,h^*=1,\dots,H}, \quad (27)$$

where each  $\mathbf{Q}_{hh^*}$  is a square matrix of order  $N_0$ , such that each element of  $\mathbf{Q}_{hh}$  is  $1/N^2$ , whereas for  $h \neq h^*$ , each element of  $\mathbf{Q}_{hh^*}$  is  $-1/(N^2(H-1))$ . Then  $\mathbf{Q}_{\text{whole-plot}} \in \mathbb{Q}$ , and condition (20) is equivalent to the constancy of  $\bar{Y}_h(z) - \bar{Y}_h(z^*)$  over  $h = 1, \dots, H$ , for every  $z, z^* \in Z$ , where  $\bar{Y}_h(z) = \sum_{i \in \Omega_h} Y_i(z)/N_0$ , (i.e., the average of all potential outcomes corresponding to treatment  $z$  in  $\Omega_h$ ) for each  $h$  and  $z$ . Hence, (20) is now equivalent to additivity of the  $\bar{Y}_h(z)$ , which can be viewed as additivity between the  $\Omega_h$  and is again milder than strict additivity; e.g., if the  $\Omega_h$  represent whole-plots as in Example 2 of Section 3.1, then (20) only requires between-whole-plot additivity, as defined by Zhao et al. (2016). Incidentally, if  $N_1, \dots, N_H$  are not all equal, then it can be impossible to find a psd counterpart of  $\mathbf{Q}_{\text{whole-plot}}$ .

Even without any split-plot type interpretation, a special case of  $\mathbf{Q}_{\text{whole-plot}}$  given by (27) can be of interest, as will be seen later in the second example of Section 5.4. For  $H = 2$  and  $N = 2N_0$ ,  $\mathbf{Q}_{\text{whole-plot}}$  is:

$$\mathbf{Q}_{\text{half}} = \begin{pmatrix} \tilde{\mathbf{Q}} & -\tilde{\mathbf{Q}} \\ -\tilde{\mathbf{Q}} & \tilde{\mathbf{Q}} \end{pmatrix} \quad (28)$$

where  $\tilde{\mathbf{Q}}$  is a square matrix of order  $N_0$  with each element equal to  $1/N^2$ . In this case, the GA condition (20) is equivalent to

$$\sum_{i=1}^{N_0} (Y_i(z) - Y_i(z^*)) = \sum_{i=N_0+1}^N (Y_i(z) - Y_i(z^*)), \quad z, z^* \in Z. \quad (29)$$

As a result,  $\mathbf{Q}_{\text{half}}$  may be a sensible choice when the finite population can be split into two groups of equal size such that each unit in the first group has a counterpart, or match, in the second group.

## 5.2 The SAP condition

Using the definitions of  $\widetilde{M}_{ii^*}(z, z^*)$  and  $B_{ii^*}(z, z^*)$  in (17) and (10) respectively, the SAP condition (22), can be stated as:

$$\pi_{ii^*}(z, z^*) = 0 \Rightarrow g(z)g(z^*) \{q_{ii^*} - (1/N^2)\} = 0. \quad (30)$$

Then a set of sufficient conditions for (30) is

$$\pi_{ii^*}(z, z^*) = 0 \text{ for at least one pair } z, z^* \in Z \Rightarrow q_{ii^*} = 1/N^2, \quad i, i^* (\neq i) = 1, \dots, N. \quad (31)$$

These conditions are also necessary if  $g(z) \neq 0$  for every  $z \in Z$ . Instead of (30), it will often help to invoke (31) because it does not involve  $g(z)$  and hence is relatively simpler. Of course, both (30) and (31) are trivially satisfied if the  $\pi_{ii^*}(z, z^*)$  are all positive, which is the case for stratified random assignment (Example 1 in Section 3.1), as can be seen from equations (40) and (41) in the Appendix.

In contrast, for the split-plot assignment (Example 2) some of the  $\pi_{ii^*}(z, z^*)$  vanish even when  $r_1(z_1) \geq 2$  for every  $z_1$ , as assumed earlier. From (42) and (43), one can see that (31) holds in this example for  $\mathbf{Q}_{\text{whole-plot}}$  in (27) with appropriate labeling of units, but not for  $\mathbf{Q}_{\text{strict}}$  in (16) or  $\mathbf{Q}_{\text{strat}}$  in (26).

Finally, in Example 3, by (44) and (45), for every  $i, i^* (\neq i)$ , either  $\pi_{ii^*}(z, z) = 0$  for all  $z$ , or  $\pi_{ii^*}(z, z^*) = 0$  for all  $z, z^* (\neq z)$ , so that (30) forces each off-diagonal element of  $\mathbf{Q}$  to be  $1/N^2$ , because any treatment contrast has  $g(z) \neq 0$  for at least two distinct  $z$ . Because no such  $\mathbf{Q}$  can be found in  $\mathbb{Q}$ , the assignment mechanism in Example 3 cannot use Theorem 1 via (21) for unbiased variance estimation, which is anticipated because uncluster sampling designs, which motivate this mechanism, are well-known in finite population sampling to have the same deficiency.

## 5.3 The strength of the generalized approach to variance estimation using the GA and SAP conditions

We now indicate how the findings in Sections 5.1 and 5.2 reveal the strength of our approach. If the strata  $\Omega_h$  in example (a) exhibit within stratum additivity, then with  $\mathbf{Q}_{\text{strat}}$  in (26), the GA condition is met. Consequently, by Theorem 2, the sampling variance of any treatment contrast estimator can be unbiasedly estimated using (21), whenever the assignment mechanism satisfies the SAP condition. This is possible for any assignment mechanism, and not just the mechanism



in Example 1 of Section 3.1, regardless of the LUEs chosen for the treatment means. Without the approach adopted here, such a general conclusion does not emerge as a naïve extension of the previous results in causal inference literature, e.g., those in Dasgupta et al. (2015) that pertain to the special case of a balanced completely randomized assignment mechanism along with mean per unit estimators, even though within stratum additivity reduces to strict additivity for  $H = 1$ . Similarly, if the  $\Omega_h$  in (b) of Section 5.1 are whole-plots and between whole-plot additivity holds, then with  $\mathbf{Q}_{\text{whole-plot}}$  in (27), the GA condition is met. Consequently, (21) can be used for unbiased variance estimation for *all* LUEs of the treatment means and under *any* assignment mechanism satisfying (31), and not just the mechanism in Example 2 of Section 3.1. It is this generality that makes Theorem 2 and the resultant estimator (21) important.

The idea of imposing a version of additivity milder than strict additivity (by choosing an appropriate matrix  $\mathbf{Q}$  that is different from  $\mathbf{Q}_{\text{strict}}$ ) to estimate the sampling variance of  $\hat{\tau}$  under less restrictive conditions appears to be attractive. However, unless there is a firm belief in the validity of the GA condition (20), the consequences of departures from it should be considered when choosing  $\mathbf{Q}$ . We discuss this aspect further in Section 6.

#### 5.4 Some examples for specific assignment mechanisms and choice of $\mathbf{Q}$

We now illustrate, how the use of HT-estimators of the treatment means along with specific choices of  $\mathbf{Q}$  that satisfy the SAP condition lead to intuitive algebraic expressions of sampling variances of treatment contrasts and their unbiased estimators when the GA condition is met. We also indicate how some of these expressions relate to existing results.

**Example 1 (stratified assignment mechanism) continued.** For each  $z$ , consider the HT-estimator defined by (7), which by (39), reduces to  $\hat{Y}^{\text{HT}}(z) = \sum_{h=1}^H N_h \hat{Y}_h(z)/N$ , where  $\hat{Y}_h(z) = \sum_{i \in T_h(z)} Y_i(z)/r_h(z)$  is the average of the observed outcomes from the  $r_h(z)$  units assigned to treatment  $z$  in stratum  $\Omega_h$ , and  $T_h(z)$  is the set of these  $r_h(z)$  units. From standard results in finite population sampling theory, for the HT-estimator,

$$\begin{aligned} A(z, z^*) &= A_i^{(1)}(z, z^*) = A_i^{(2)}(z, z^*) = 0, \\ B_{ii}(z, z) &= 1/(N^2 \pi_i(z)), \quad B_{ii^*}(z, z^*) = \pi_{ii^*}(z, z^*)/(N^2 \pi_i(z) \pi_{i^*}(z^*)). \end{aligned} \quad (32)$$

Substituting the first and second-order assignment probabilities for Example 1 (from equations

(39) to (41) in the Appendix) into (32), gives, for any  $i \in \Omega_h$ ,  $i^* (\neq i) \in \Omega_{h^*}$ ,

$$\begin{aligned} B_{ii}(z, z) &= \frac{N_h}{N^2 r_h(z)}, \quad B_{ii^*}(z, z) = \frac{N_h (r_h(z) - 1)}{N^2 (N_h - 1) r_h(z)}, \quad \text{if } h = h^* \\ B_{ii^*}(z, z^*) &= \frac{N_h}{N^2 (N_h - 1)} \quad \text{if } h = h^*, z \neq z^*, \quad B_{ii^*}(z, z^*) = \frac{1}{N^2} \quad \text{if } h \neq h^*. \end{aligned} \quad (33)$$

Now, label the units so that  $\Omega_1 = \{1, \dots, N_1\}$ ,  $\Omega_2 = \{N_1 + 1, \dots, N_1 + N_2\}$ , etc. Then with  $\mathbf{Q} = \mathbf{Q}_{\text{strat}}$  as in (26), substitution of (33) into (17) yields

$$\widetilde{M}_{ii^*}(z, z^*) = \begin{cases} -\frac{(g(z))^2 N_h}{N^2 (N_h - 1) r_h(z)}, & \text{if } z = z^* \text{ and } i, i^* \in \Omega_h \text{ for some } h \\ 0 & \text{otherwise.} \end{cases}$$

Similar expressions for  $M, M_i(z), M_{ii}(z)$  can be obtained by substituting (33) into (12). Finally, substituting the expressions of  $M, M_i(z), M_{ii}(z)$  and  $\widetilde{M}_{ii^*}(z, z^*)$  into (18) and (21), and after some simplification, we obtain:

$$V_{\mathbf{Q}_{\text{strat}}}(\widehat{\tau}^{\text{HT}}) = \frac{1}{N^2} \sum_{z \in Z} \sum_{h=1}^H (g(z))^2 \frac{N_h^2}{r_h(z)} S_h(z, z), \quad (34)$$

$$\hat{V}_{\mathbf{Q}_{\text{strat}}}(\widehat{\tau}^{\text{HT}}) = \frac{1}{N^2} \sum_{z \in Z} \sum_{h=1}^H (g(z))^2 \frac{N_h^2}{r_h(z)} \hat{S}_h(z, z), \quad (35)$$

where  $S_h(z, z) = \sum_{i \in \Omega_h} (Y_i(z) - \bar{Y}_h(z))^2 / (N_h - 1)$  is the variance (with divisor  $N_h - 1$ ) and  $\bar{Y}_h(z) = \sum_{i \in \Omega_h} Y_i(z) / N_h$  is the average of all potential outcomes of units in stratum  $\Omega_h$  for treatment  $z$ , and

$$\hat{S}_h(z, z) = \sum_{i \in T_h(z)} \left( Y_i(z) - \widehat{Y}_h(z) \right)^2 / (r_h(z) - 1)$$

is an unbiased estimator of  $S_h(z, z)$ , calculated as the sample variance of observed outcomes in stratum  $\Omega_h$  that correspond to treatment  $z$ .

Also  $\boldsymbol{\tau}' \mathbf{Q}_{\text{strat}} \boldsymbol{\tau} = \sum_{h=1}^H N_h S_h(\tau, \tau) / N^2$ , where  $S_h(\tau, \tau) = \sum_{i \in \Omega_h} (\tau_i - \bar{\tau}^{(h)})^2 / (N_h - 1)$  is the variance (with divisor  $N_h - 1$ ) and  $\bar{\tau}^{(h)} = \sum_{i \in \Omega_h} \tau_i / N_h$  is the average of the unit-level treatment contrasts within stratum  $\Omega_h$ . Consequently, by Theorem 2, part (1), we obtain

$$\text{var}(\widehat{\tau}^{\text{HT}}) = \frac{1}{N^2} \sum_{h=1}^H N_h \left( \sum_{z \in Z} (g(z))^2 \frac{N_h}{r_h(z)} S_h(z, z) - S_h(\tau, \tau) \right), \quad (36)$$

When  $H = 1$ , a stratified assignment mechanism reduces to a completely randomized assignment, and  $\mathbf{Q}_{\text{strat}}$  reduces to  $\mathbf{Q}_{\text{strict}}$ . Consequently, expressions (34), (35) and (36) reduce to the expressions derived by Dasgupta et al. (2015). When  $Z = \{0, 1\}$ , (36) further reduces to the Neymanian decomposition (13) discussed in Section 4.1.

**Example on completely randomized assignment using  $\mathbf{Q}_{\text{half}}$ .** Let  $N = 2N_0$  so that  $\mathbf{Q}_{\text{half}}$  in (28) is well-defined. Consider a completely randomized assignment that allocates  $r(z)$  ( $\geq 2$ ),  $z \in Z$ , units to treatment  $z$ , where the  $r(z)$ 's sum to  $N$ . As in the previous example, for each  $z$ , consider the HT-estimator  $\hat{Y}^{\text{HT}}(z) = \sum_{i \in T(z)} Y_i(z)/r(z)$ . Because the assignment mechanism in this example is a special case of the previous example with  $H = 1$ , dropping the index  $h$  from the expressions in (33) results in

$$\begin{aligned} B_{ii}(z, z) &= \frac{1}{Nr(z)}, \quad B_{ii^*}(z, z) = \frac{r(z) - 1}{N(N-1)r(z)}, \\ B_{ii^*}(z, z^*) &= \frac{1}{N(N-1)} \text{ if } z \neq z^*. \end{aligned}$$

Substituting the above expressions into (12) and (17), and further substituting the resultant expressions of  $M$ 's and  $\widetilde{M}$ 's into (18) and (21), after considerable algebra, we obtain

$$\begin{aligned} V_{\mathbf{Q}_{\text{half}}}(\hat{\tau}^{\text{HT}}) &= \sum_{z \in Z} \frac{(g(z))^2}{r(z)} S(z, z) + \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{i^* (\neq i)=1}^N \tau_i \tau_{i^*} - \bar{\tau}^{(1)} \bar{\tau}^{(2)}, \\ \hat{V}_{\mathbf{Q}_{\text{half}}}(\hat{\tau}^{\text{HT}}) &= \hat{\tau}^2 - \frac{4(N-1)}{N} \sum_{z \in Z} \frac{(g(z))^2}{r(z)(r(z)-1)} L_1(z) L_2(z) \\ &\quad - \frac{4(N-1)}{N} \sum_{z \in Z} \sum_{z^* (\neq z) \in Z} \frac{g(z)g(z^*)}{r(z)r(z^*)} L_1(z) L_2(z^*), \end{aligned}$$

where

$$S(z, z) = \sum_{i=1}^N (Y_i(z) - \bar{Y}(z))^2 / (N-1)$$

is the variance (with divisor  $N-1$ ) of potential outcomes of all units for treatment  $z$ ,

$$\bar{\tau}^{(1)} = \frac{1}{N_0} \sum_{i=1}^{N_0} \tau_i \quad \text{and} \quad \bar{\tau}^{(2)} = \frac{1}{N_0} \sum_{i=N_0+1}^N \tau_i$$

are the averages of unit-level estimands of units  $\{1, \dots, N_0\}$  and  $\{N_0 + 1, \dots, N\}$  respectively, and  $L_1(z)$  and  $L_2(z)$  are the sums of observed outcomes of units, respectively in  $\{1, \dots, N_0\}$  and  $\{N_0 + 1, \dots, N\}$ , that are exposed to treatment  $z$ .

The estimator  $\hat{V}_{\mathbf{Q}_{\text{half}}}(\hat{\tau}^{\text{HT}})$  unbiasedly estimates  $\text{var}(\hat{\tau}^{\text{HT}})$  if and only if (29) holds. Also, note that  $\boldsymbol{\tau}' \mathbf{Q}_{\text{half}} \boldsymbol{\tau} = (\bar{\tau}^{(1)} - \bar{\tau}^{(2)})^2 / 4$ . Therefore, by Theorem 2, part (1),

$$\text{var}(\hat{\tau}^{\text{HT}}) = V_{\mathbf{Q}_{\text{half}}}(\hat{\tau}^{\text{HT}}) - \boldsymbol{\tau}' \mathbf{Q}_{\text{half}} \boldsymbol{\tau} = \sum_{z \in Z} \frac{(g(z))^2}{r(z)} S(z, z) - \frac{1}{N(N-1)} \sum_{i=1}^N (\tau_i - \bar{\tau})^2,$$

which is again in agreement with the results derived in Dasgupta et al. (2015), who decomposed  $\text{var}(\hat{\tau}^{\text{HT}})$  using  $\mathbf{Q}_{\text{strict}}$  rather than  $\mathbf{Q}_{\text{half}}$ .

## 6 More on additivity: Is the Neymanian strategy “optimal”?

With a known assignment mechanism, if there are strong reasons to believe that there exists a psd  $\mathbf{Q} \in \mathbb{Q}$  satisfying the SAP and GA conditions, then the results in Section 4.2 suggest that one can use the  $\hat{V}_{\mathbf{Q}}(\hat{\tau})$ , obtained from (17) and (21), to unbiasedly estimate  $var(\hat{\tau})$  for any treatment contrast  $\bar{\tau}$ . For instance, with the stratified assignment mechanism in Example 1, if one is confident about within stratum additivity, then as discussed in (a) of Section 5.1,  $\mathbf{Q}_{\text{strat}}$  in (26) can be used for unbiased variance estimation.

The GA condition (20), however, involves the potential outcomes, at least half of which are not observed and, as a result, there can be situations where one is not sure about the validity of such condition for any  $\mathbf{Q}$ . Then, in order to choose  $\mathbf{Q}$ , one should consider the consequences of departures from (20) in terms of the ensuing bias in variance estimation, with a view to limiting it.

### 6.1 Positive second-order assignment probabilities

We begin with assignment mechanisms for which the second-order assignment probabilities are all positive, a property in the survey literature that describes a measurable survey design (Kish, 1965). A simplifying feature then is that the SAP condition is irrelevant, because it is trivially met by every  $\mathbf{Q} \in \mathbb{Q}$ . Consider such a psd matrix  $\mathbf{Q}$ . By Theorem 2, part (1), use of the resulting  $\hat{V}_{\mathbf{Q}}(\hat{\tau})$  as an estimator of  $var(\hat{\tau})$  will lead to a non-negative bias of magnitude  $\boldsymbol{\tau}'\mathbf{Q}\boldsymbol{\tau}$  in the absence of any knowledge about the GA condition (20). To understand the implications of choosing such a  $\mathbf{Q}$  and the resultant estimator  $\hat{V}_{\mathbf{Q}}(\hat{\tau})$ , consider three scenarios: (i) strict additivity holds (which implies that all milder versions of additivity also hold), (ii) strict additivity does not hold, but the GA condition (20) holds for some  $\mathbf{Q}^* \in \mathbb{Q}$  (e.g.,  $\mathbf{Q}_{\text{strat}}$ ) (iii) Neither  $\mathbf{Q}_{\text{strict}}$  nor  $\mathbf{Q}^* \in \mathbb{Q}$  satisfies the GA condition (20). We consider two possible estimators of  $var(\hat{\tau})$ : First, the generalized Neymanian variance estimator  $\hat{V}_{\mathbf{Q}_{\text{strict}}}(\hat{\tau})$ , and second,  $\hat{V}_{\mathbf{Q}^*}(\hat{\tau})$  based on  $\mathbf{Q}^*$ . Table 1 shows the bias of estimation associated with these situations and two choices for estimators.

Table 1 shows that if strict additivity holds, then both estimators are unbiased. If milder additivity holds, but not strict additivity, then the Neymanian estimator has positive bias whereas the one based on  $\mathbf{Q}^*$  is unbiased. However, if even milder additivity does not hold, then both sampling variance estimators are biased, and none of them has a larger bias than the other uniformly in  $\boldsymbol{\tau}$ . This is because,  $\mathbf{Q}_{\text{strict}}$  and  $\mathbf{Q}^*$  have all diagonal elements  $1/N^2$ , and hence unless

Table 1: Bias in the estimation of  $\text{var}(\hat{\tau})$  under six different scenarios

Truth	Variance Estimator	
	$\hat{V}_{\mathbf{Q}_{\text{strict}}}(\hat{\tau})$ (Neyman)	$\hat{V}_{\mathbf{Q}^*}(\hat{\tau})$
$\mathbf{Q}_{\text{strict}}$ satisfies (20) (strict additivity)	0	0
$\mathbf{Q}^*$ satisfies (20) but not $\mathbf{Q}_{\text{strict}}$	$\boldsymbol{\tau}'\mathbf{Q}_{\text{strict}}\boldsymbol{\tau}$	0
Neither $\mathbf{Q}_{\text{strict}}$ nor $\mathbf{Q}^*$ satisfies (20)	$\boldsymbol{\tau}'\mathbf{Q}_{\text{strict}}\boldsymbol{\tau}$	$\boldsymbol{\tau}'\mathbf{Q}^*\boldsymbol{\tau}$

they are equal, their difference is neither positive semidefinite nor negative semidefinite. Thus the issue of bias control becomes nontrivial, and appropriate choice of  $\mathbf{Q} \in \mathbb{Q}$  from this perspective becomes important.

Consider minimizing the maximum possible bias to protect against the worst possible scenario, which arises when the bias  $\boldsymbol{\tau}'\mathbf{Q}\boldsymbol{\tau}$  is maximum with respect to  $\boldsymbol{\tau}$ . Whereas for any  $\mathbf{Q}$ , the set of values of  $\boldsymbol{\tau}'\mathbf{Q}\boldsymbol{\tau}$  over all possible  $\boldsymbol{\tau}$  is not bounded above, it is also true that  $\boldsymbol{\tau}$  essentially represents the same vector of unit-level treatment contrasts even after multiplication by any nonzero constant. Hence, in the spirit of E-optimality in experimental design (Ehrenfeld, 1955), consider choosing  $\mathbf{Q} \in \mathbb{Q}$  to minimize

$$\max_{\boldsymbol{\tau}: \boldsymbol{\tau}'\boldsymbol{\tau}=1} \boldsymbol{\tau}'\mathbf{Q}\boldsymbol{\tau} = \lambda_{\max}(\mathbf{Q}), \quad (37)$$

where  $\lambda_{\max}(\mathbf{Q})$  is the largest eigenvalue of  $\mathbf{Q}$ . The following proposition (proof in Appendix) provides a guideline for choosing the optimal  $\mathbf{Q}$  with respect to criterion (37)

**Proposition 1.** *For any psd matrix  $\mathbf{Q} \in \mathbb{Q}$ ,*

$$\lambda_{\max}(\mathbf{Q}) \geq \frac{1}{N(N-1)},$$

*with equality if and only if  $\mathbf{Q} = \mathbf{Q}_{\text{strict}}$ .*

For  $\mathbf{Q} = \mathbf{Q}_{\text{strict}}$ , recall that the GA condition reduces to the strict additivity condition. Thus, if the second-order assignment probabilities are all positive, the choice of  $\mathbf{Q}$  that makes the GA condition most demanding is also most protective against the worst scenario regarding bias.

*Remark 4.* Proposition 1 provides a justification for using the Neymanian estimator of  $\text{var}(\hat{\tau})$ : it minimizes the maximum bias of sampling variance estimation *for any assignment mechanism* for which all second-order assignment probabilities are positive.

## 6.2 Case of split-plot assignment mechanism

Assignment mechanisms with some vanishing second-order assignment probabilities rely on the SAP condition, which may impose messy patterns on  $\mathbf{Q}$ . Indeed, as with the assignment mechanism in Example 3 of Section 3.1, there may not even exist any  $\mathbf{Q} \in \mathbb{Q}$  satisfying the SAP condition. The split-plot assignment mechanism in Example 2, however, allows an extension of the ideas in Section 6.1.

**Proposition 2.** *Let the  $N = HN_0$  units be labeled such that the  $H$  whole-plots are taken as  $\Omega_1 = \{1, \dots, N_0\}, \dots, \Omega_H = \{(H-1)N_0 + 1, \dots, HN_0\}$ .*

- (a) *For the split-plot assignment mechanism, a psd matrix  $\mathbf{Q} \in \mathbb{Q}$  satisfies the SAP condition (31) if and only if it is of the form  $\mathbf{Q} = \mathbf{Q}_1 \otimes (\mathbf{1}_0 \mathbf{1}_0')$ , where  $\otimes$  denotes Kronecker product,  $\mathbf{Q}_1$  is a psd matrix of order  $H$  having each row sum zero and each diagonal element  $1/N^2$ , and  $\mathbf{1}_0$  is the  $N_0 \times 1$  vector of all ones.*
- (b) *For any psd matrix  $\mathbf{Q}$  as in (a) above,*

$$\lambda_{\max}(\mathbf{Q}) \geq \frac{1}{N(H-1)},$$

*with equality if and only if  $\mathbf{Q}_1$  has each diagonal element  $1/N^2$  and each off-diagonal element  $-1/(N^2(H-1))$ .*

The proof of Proposition 2 is in the Appendix. For the split-plot assignment mechanism, one needs to consider  $\mathbf{Q}$  as in Proposition 2(a). With  $\mathbf{Q}_1$  as in Proposition 2(b), such a  $\mathbf{Q}$  equals  $\mathbf{Q}_{\text{whole-plot}}$ , the one shown in (27). Hence,  $\mathbf{Q}_{\text{whole-plot}}$  is uniquely the best choice of  $\mathbf{Q} \in \mathbb{Q}$ , for minimizing the maximum possible bias given by (37).

## 7 Treatments with a factorial structure

The ideas discussed apply to situations where the treatments have a factorial structure. To see this, consider an  $s_1 \times \dots \times s_K$  factorial experiment with  $K$  factors  $F_k$ ,  $k = 1, \dots, K$ , with the levels of  $F_k$  denoted as  $0, 1, \dots, s_{k-1}$ . As earlier, the treatment combinations are  $z \in Z$ , and let  $\bar{Y}(z) = \sum_{i=1}^N Y_i(z)/N$ , where  $Y_i(z)$  is the potential outcome of unit  $i$  when exposed to treatment combination  $z$ . Let  $\bar{\mathbf{Y}}$  denote the column vector of order  $|Z|$  having elements  $\bar{Y}(z)$ , arranged

lexicographically, e.g., in a  $2 \times 3$  factorial experiment,

$$\bar{\mathbf{Y}} = (\bar{Y}(00), \bar{Y}(01), \bar{Y}(02), \bar{Y}(10), \bar{Y}(11), \bar{Y}(12))'.$$

Then a typical treatment contrast belonging to a factorial effect  $F_1^{x_1} \dots F_K^{x_K}$ , where  $x_1, \dots, x_K$  are binary and at least one is non-zero, is of the form  $\mathbf{g}'\bar{\mathbf{Y}}$ , with  $\mathbf{g}$  having the structure  $\mathbf{g} = \mathbf{g}_1 \otimes \dots \otimes \mathbf{g}_K$ , such that each  $\mathbf{g}_k$  is a nonnull vector having all elements equal if  $x_k = 0$ , and elements summing to zero if  $x_k = 1$ , ( $k = 1, \dots, K$ ) (see, for example, Chapter 2 of Gupta and Mukerjee (1989)). A contrast of this kind is of the form (2), with  $g(z)$  representing the element of  $\mathbf{g}$  that corresponds to treatment combination  $z$ . Thus, all our results are applicable to treatment contrasts representing factorial main effects or interactions even under a general factorial structure for the treatments.

## 8 Simulations and numerical examples

Some simulations demonstrate the usefulness of the proposed approach and the effects of departures from strict and mild additivity on biases of proposed sampling variance estimators. Consider a hypothetical randomized experiment with  $N = 50$  units, designed to estimate the causal effects of three treatments labeled 1, 2 and 3. Assume that interest lies in the contrast  $\bar{\tau} = \bar{Y}(1) - 2\bar{Y}(2) + \bar{Y}(3)$ , where  $\bar{Y}(z)$  denotes the mean of the 50 potential outcomes for treatment  $z$ ,  $z \in \{1, 2, 3\}$ . Such a contrast is often of practical interest: for qualitative treatments it can be interpreted as the causal effect of treatment 2 compared to the average effect of treatments 1 and 3, and for quantitative treatments (e.g., three levels of temperature) it represents the quadratic effect of the treatments on the outcome.

We also assume that a blocking factor (e.g., sex) is used to stratify the population of 50 units into two strata  $\Omega_1$  and  $\Omega_2$  of sizes 30 and 20, respectively. Accordingly, we define a finite population as a  $50 \times 3$  matrix of potential outcomes, whose  $i$ th row is denoted by  $\mathbf{Y}_i = (Y_i(1), Y_i(2), Y_i(3))$ ,  $i = 1, \dots, 50$ , where each  $\mathbf{Y}_i$  is generated in our simulation using the trivariate normal model:

$$\mathbf{Y}_i \sim N_3(\boldsymbol{\mu}_h, \boldsymbol{\Sigma}_h), \quad i \in \Omega_h, \quad h = 1, 2,$$

where

$$\boldsymbol{\Sigma}_h = \sigma_h^2 \begin{pmatrix} 1 & \rho_h & \rho_h \\ \rho_h & 1 & \rho_h \\ \rho_h & \rho_h & 1 \end{pmatrix}, \quad h = 1, 2.$$

The choices of  $\mu_h, \sigma_h^2$  and  $\rho_h$  for  $h = 1, 2$  determine the level of additivity in populations generated from the above model. We study the biases of the estimators of  $var(\hat{\tau})$  for six such generating models with different levels of additivity. These models are shown in Table 2. Populations generated from models I to VI are in increasing order of departure from strict additivity. Note that strict additivity holds for model I and stratum-level additivity holds for model II. Neither form of additivity holds for models III-VI; specifically, for model V, potential outcomes for the three treatments are generated from independent populations, and for model VI, potential outcomes for each pair of treatments have a negative correlation of -0.5.

Table 2: Parameters of models used to generate simulated potential outcomes

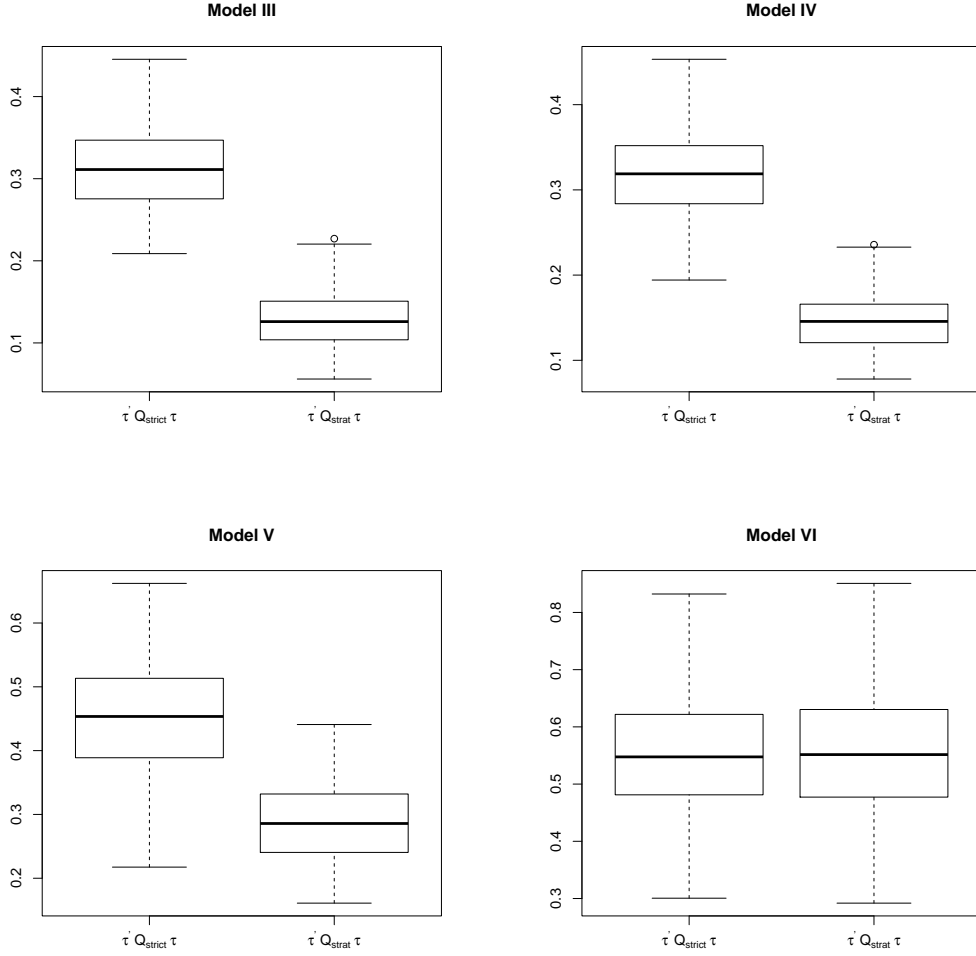
Model	Generating parameters				
	$\mu$ 's	$\sigma_1^2$	$\sigma_2^2$	$\rho_1$	$\rho_2$
I	$\mu_1 = \mu_2 = (8, 7, 10)$	2	2	1	1
II	$\mu_1 = (10, 12, 14), \mu_2 = (8, 6, 10)$	2	3	1	1
III	$\mu_1 = (10, 12, 14), \mu_2 = (8, 6, 10)$	2	3	.2	.9
IV	$\mu_1 = (10, 12, 14), \mu_2 = (8, 6, 10)$	2	3	.5	.5
V	$\mu_1 = (10, 12, 14), \mu_2 = (8, 6, 10)$	2	3	0	0
VI	$\mu_1 = \mu_2 = (8, 7, 10)$	3	3	-.5	-.5

For each population, we consider a stratified assignment mechanism, which, as noted earlier, meets the SAP condition for every choice of  $Q$ , and compare two estimators of  $var(\hat{\tau})$  obtained by substituting  $Q_{\text{strict}}$  from (16) and  $Q_{\text{strat}}$  from (26) for  $Q$  in  $\hat{V}_Q(\hat{\tau})$  given by (21). By Theorem 2, these sampling variance estimators have biases  $\tau'Q_{\text{strict}}\tau$  and  $\tau'Q_{\text{strat}}\tau$ , respectively. We generate 100 populations from each of the models I–VI, and for each population (and hence each vector  $\tau$  of unit-level estimands) generated from a specific model, compute the biases  $\tau'Q_{\text{strict}}\tau$  and  $\tau'Q_{\text{strat}}\tau$ , along with their ratio  $\tau'Q_{\text{strat}}\tau/\tau'Q_{\text{strict}}\tau$ .

The results are seen to follow a pattern consistent with our results. In populations generated from model I, where strict additivity holds, both sampling variance estimators lead to zero bias for each  $\tau$ . For model II, where within stratum additivity holds but not strict additivity,  $\tau'Q_{\text{strat}}\tau$  vanishes for every  $\tau$ , but using  $Q_{\text{strict}}$  results in a small positive median bias of 0.176. The summaries (boxplots) of biases for populations generated from models III, IV, V and VI, for which neither form of additivity holds, are shown in Figure 1. The medians of the ratios of the



Figure 1: Boxplots of bias with  $\mathbf{Q}_{\text{strict}}$  and  $\mathbf{Q}_{\text{strat}}$  for populations generated from models III-VI



biases are

$$\text{median}_{\tau} \left( \frac{\tau' \mathbf{Q}_{\text{strat}} \tau}{\tau' \mathbf{Q}_{\text{strict}} \tau} \right) = 0.42, 0.46, 0.63 \text{ and } 1.01, \quad (38)$$

for models III, IV, V and VI respectively. By (38), the biases arising from use of  $\mathbf{Q}_{\text{strat}}$  are substantially smaller than those arising from use of  $\mathbf{Q}_{\text{strict}}$ , except in model VI, which entails negative correlation among pairwise columns of potential outcomes, and the largest bias among all cases, partially justifying  $\mathbf{Q}_{\text{strict}}$  as a minimax solution. Because negatively correlated potential outcomes are unlikely to arise in practice, choosing  $\mathbf{Q}_{\text{strict}}$  as a minimax solution can, however, be a bit pessimistic in the context of this example. A criterion of average bias, as indicated in Section 9 is an alternative choice.

## 9 Discussion

This article extends the existing theory of causal inference for treatment contrasts from randomized experiments in several ways and sheds light on the existing results from different perspectives. First, it develops a Neymanian inferential procedure for a *general assignment mechanism* that includes, as special cases, most of the commonly used randomization and restricted randomization procedures for assigning experimental units to treatments. Second, it integrates statistical inference for causal effects with statistical inference for finite population estimands. Third, it considers multiple treatments allowing the possibility of symmetric or asymmetric factorial structure. Fourth, it proposes conditions that are milder than the well-known strict additivity condition for unbiased estimation of the sampling variance of any treatment contrast estimator. Finally, a new justification for using the Neymanian conservative sampling variance estimator is obtained. Two aspects that may be of additional interest, but involve a *superpopulation perspective* rather than the finite population approach adopted in this paper, are briefly indicated below.

The first of these concerns a criterion of average or expected bias, rather than that of maximum possible bias considered in Section 6, the expectation being with respect to a prior distribution on  $\boldsymbol{\tau}$ , as induced by a superpopulation specification on the potential outcomes. This calls for choosing  $\mathbf{Q} \in \mathbb{Q}$  so as to minimize  $E(\boldsymbol{\tau}'\mathbf{Q}\boldsymbol{\tau}) = \text{tr}(\mathbf{Q}\boldsymbol{\Gamma})$ , where  $\boldsymbol{\Gamma} = E(\boldsymbol{\tau}\boldsymbol{\tau}')$  and the expectation is defined with respect to the superpopulation model. Although this optimization problem does not seem to admit an analytical solution in general, preliminary studies suggest that a semidefinite programming formulation should help. In contrast to Proposition 1, even if the second-order assignment probabilities are all positive, the resulting optimal  $\mathbf{Q}$ , now dependent on  $\boldsymbol{\Gamma}$ , can be quite different from  $\mathbf{Q}_{\text{strict}}$ .

From the perspective of a superpopulation model, it is also of interest to explore an optimal inference strategy, i.e., an optimal combination of an assignment mechanism and LUEs  $\hat{Y}(z)$ ,  $z \in Z$ , that minimizes the model-expected total sampling variance of estimators of treatment contrasts of interest. Our general framework would allow borrowing ideas from finite population sampling for this purpose.

Work is currently in progress on the problems mentioned above and will be reported elsewhere.

## Appendix

### Assignment probabilities in Example 1:

For  $i \in \Omega_h$ ,  $h = 1, \dots, H$ , and  $z \in Z$ ,

$$\pi_i(z) = r_h(z)/N_h. \quad (39)$$

For  $i, i^* (\neq i) \in \Omega_h$ ,  $h = 1, \dots, H$ , and  $z, z^* \in Z$ ,

$$\begin{aligned} \pi_{ii^*}(z, z) &= \frac{r_h(z)(r_h(z) - 1)}{N_h(N_h - 1)}, \\ \pi_{ii^*}(z, z^*) &= \frac{r_h(z)r_h(z^*)}{N_h(N_h - 1)}, \quad z \neq z^*. \end{aligned} \quad (40)$$

For  $i \in \Omega_h, i^* \in \Omega_{h^*}$ ,  $h, h^* (\neq h) = 1, \dots, H$ , and  $z, z^* \in Z$ ,

$$\pi_{ii^*}(z, z^*) = \frac{r_h(z)r_{h^*}(z^*)}{N_h N_{h^*}}. \quad (41)$$

### Assignment probabilities in Example 2:

For  $i, i^* (\neq i) \in \Omega_h$  and  $h = 1, \dots, H$ ,

$$\begin{aligned} \pi_i(z_1 z_2) &= \frac{r_1(z_1)r_2(z_2)}{N}, \quad z_1 z_2 \in Z, \\ \pi_{ii^*}(z_1 z_2, z_1 z_2) &= \frac{r_1(z_1)r_2(z_2)(r_2(z_2) - 1)}{N(N_0 - 1)}, \quad z_1 z_2 \in Z, \\ \pi_{ii^*}(z_1 z_2, z_1 z_2^*) &= \frac{r_1(z_1)r_2(z_2)r_2(z_2^*)}{N(N_0 - 1)}, \quad z_1 z_2, z_1 z_2^* (z_2 \neq z_2^*) \in Z, \\ \pi_{ii^*}(z_1 z_2, z_1^* z_2^*) &= 0, \quad z_1 z_2, z_1^* z_2^* (z_1 \neq z_1^*) \in Z. \end{aligned} \quad (42)$$

For  $i \in \Omega_h, i^* (\neq i) \in \Omega_{h^*}$  and  $h, h^* (\neq h) = 1, \dots, H$ ,

$$\begin{aligned} \pi_{ii^*}(z_1 z_2, z_1 z_2^*) &= \frac{r_1(z_1)(r_1(z_1) - 1)r_2(z_2)r_2(z_2^*)}{N N_0 (H - 1)}, \quad z_1 z_2, z_1 z_2^* \in Z, \\ \pi_{ii^*}(z_1 z_2, z_1^* z_2^*) &= \frac{r_1(z_1)r_1(z_1^*)r_2(z_2)r_2(z_2^*)}{N N_0 (H - 1)}, \quad z_1 z_2, z_1^* z_2^* (z_1 \neq z_1^*) \in Z. \end{aligned} \quad (43)$$

### Assignment probabilities in Example 3:

For  $i, i^* (\neq i) \in \Delta(l)$  and  $l = 1, \dots, |Z|$ ,

$$\begin{aligned} \pi_i(z) &= 1/|Z|, \quad z \in Z, \\ \pi_{ii^*}(z, z) &= 1/|Z|, \quad z \in Z, \\ \pi_{ii^*}(z, z^*) &= 0, \quad z, z^* (\neq z) \in Z, \end{aligned} \quad (44)$$

while for  $i \in \Delta(l), i^* \in \Delta(l^*)$  and  $l, l^* (\neq l) = 1, \dots, |Z|$ ,

$$\begin{aligned}\pi_{ii^*}(z, z) &= 0, \quad z \in Z, \\ \pi_{ii^*}(z, z^*) &= \frac{1}{|Z|(|Z| - 1)}, \quad z, z^* (\neq z) \in Z.\end{aligned}\tag{45}$$

### Proof of Theorem 2:

For any  $\mathbf{Q} \in \mathbb{Q}$ , from (2), we have that

$$\begin{aligned}\bar{\tau}^2 &= \frac{1}{N^2} \boldsymbol{\tau}' \mathbf{J} \boldsymbol{\tau} = \boldsymbol{\tau}' \mathbf{Q} \boldsymbol{\tau} - \boldsymbol{\tau}' (\mathbf{Q} - (1/N^2) \mathbf{J}) \boldsymbol{\tau} \\ &= \boldsymbol{\tau}' \mathbf{Q} \boldsymbol{\tau} - \sum_{i=1}^N \sum_{i^* (\neq i)=1}^N (q_{ii^*} - 1/N^2) \tau_i \tau_{i^*}, \text{ by the second equation in (15)} \\ &= \boldsymbol{\tau}' \mathbf{Q} \boldsymbol{\tau} - \sum_{z \in Z} \sum_{z^* \in Z} \sum_{i=1}^N \sum_{i^* (\neq i)=1}^N (q_{ii^*} - 1/N^2) g(z) g(z^*) Y_i(z) Y_{i^*}(z^*), \text{ by (1)} \\ &= \boldsymbol{\tau}' \mathbf{Q} \boldsymbol{\tau} + \sum_{z \in Z} \sum_{z^* \in Z} \sum_{i=1}^N \sum_{i^* (\neq i)=1}^N \left( M_{ii^*}(z, z^*) - \widetilde{M}_{ii^*}(z, z^*) \right) Y_i(z) Y_{i^*}(z^*), \text{ by (12) and (17)}.\end{aligned}$$

Substituting the expression above for  $\bar{\tau}^2$  in (11) and applying (18), we get

$$\text{var}(\widehat{\tau}) = V_{\mathbf{Q}}(\widehat{\tau}) - \boldsymbol{\tau}' \mathbf{Q} \boldsymbol{\tau}.$$

This proves part (1) of the Theorem. Part (2) follows from the fact that  $\mathbf{Q}$  is psd by definition. To prove part (3), we note that because  $\mathbf{Q}$  is psd,  $\boldsymbol{\tau}' \mathbf{Q} \boldsymbol{\tau} = 0$  if and only if  $\mathbf{Q} \boldsymbol{\tau} = 0$ . From (1),  $\boldsymbol{\tau} = \sum_{z \in Z} g(z) \mathbf{Y}(z)$ . Thus, the upper bound in part (2) is attained for every treatment contrast if and only if  $\mathbf{Q} \mathbf{Y}(z)$  is the same for each  $z \in Z$ , or equivalently, if and only if (20) holds.

### Proof of Proposition 1

By (15),  $\mathbf{Q}$  has at least one eigenvalue equal to zero and trace  $1/N$ . Hence the mean of the positive eigenvalues of  $\mathbf{Q}$  is at least  $1/(N(N-1))$ . Because  $\lambda_{\max}(\mathbf{Q})$  is at least as large as this mean, it must satisfy

$$\lambda_{\max}(\mathbf{Q}) \geq \frac{1}{N(N-1)}.$$

Equality is attained if and only if  $\mathbf{Q}$  has  $N-1$  eigenvalues equal to  $1/(N(N-1))$ . There exists only one such  $\mathbf{Q} \in \mathbb{Q}$ , which is  $\mathbf{Q}_{\text{strict}}$ .

### Proof of Proposition 2

- (a) By equations (42) and (43) in the Appendix, for the split-plot assignment mechanism, a psd matrix  $\mathbf{Q} \in \mathbb{Q}$  satisfies (31) if and only if  $q_{ii^*} = 1/N^2$ , for every  $i, i^* \in \Omega_h$  and every  $h = 1, \dots, H$ . Hence the if part is evident. To prove the only if part, observe that any psd matrix  $\mathbf{Q} \in \mathbb{Q}$  can be expressed as  $\mathbf{Q} = \mathbf{Q}_0 \mathbf{Q}'_0$  for some  $\mathbf{Q}_0$ . Partition  $\mathbf{Q}_0$  as  $\mathbf{Q}_0 = [\mathbf{Q}'_{01}, \dots, \mathbf{Q}'_{0H}]'$ , where  $\mathbf{Q}_{0h}$  has  $N_0$  rows,  $h = 1, \dots, H$ . Suppose  $\mathbf{Q}$  satisfies (31). Then, as noted above, every diagonal block of  $\mathbf{Q}$  has each element  $1/N^2$ , i.e.,  $\mathbf{Q}_{0h} \mathbf{Q}'_{0h} = (1/N^2) (\mathbf{1}_0 \mathbf{1}'_0)$  for every  $h$ . Because the column space of  $\mathbf{Q}_{0h}$  equals that of  $\mathbf{Q}_{0h} \mathbf{Q}'_{0h}$ , it follows that, for every  $h$ ,  $\mathbf{Q}_{0h}$  has all elements equal in each column. Hence  $\mathbf{Q} = \mathbf{Q}_0 \mathbf{Q}'_0$  must be of the form  $\mathbf{Q}_1 \otimes (\mathbf{1}_0 \mathbf{1}'_0)$ , for some square matrix  $\mathbf{Q}_1$  of order  $H$ . The matrix  $\mathbf{Q}_1$  has to be psd with each row sum zero and each diagonal element  $1/N^2$ , as  $\mathbf{Q}$  is psd and  $\mathbf{Q} \in \mathbb{Q}$ .
- (b) Any psd  $\mathbf{Q}$  as in (a) has trace  $1/N$  and at most  $H - 1$  positive eigenvalues. Hence the mean of the positive eigenvalues of  $\mathbf{Q}$  is at least  $1/(N(H - 1))$ . The lower bound on  $\lambda_{\max}(\mathbf{Q})$  is now immediate as in Proposition 1. This bound is attained if and only if  $\mathbf{Q}$  has  $H - 1$  eigenvalues equal to  $1/(N(H - 1))$ , i.e., if and only if  $\mathbf{Q}_1$  has  $H - 1$  eigenvalues equal to  $1/(NN_0(H - 1))$ . As  $N = HN_0$  and each row sum of  $\mathbf{Q}_1$  is zero, this happens if and only if  $\mathbf{Q}_1$  has each diagonal element  $1/N^2$  and each off-diagonal element  $-1/(N^2(H - 1))$ .

## References

- Chakraborty, B., Collins, L. M., Strecher, V. J., and Murphy, S. A. (2009), “Developing multicomponent interventions using fractional factorial designs,” *Statistics in Medicine*, 28, 2687–2708.
- Chaudhuri, A. (2014), *Modern Survey Sampling*, CRC Press: Boca Raton, FL.
- Cochran, W. G. (1977), *Sampling Techniques (3rd ed.)*, Wiley: New York, NY.
- Collins, L. M., Dziak, J. J., and Li, R. (2009), “Design of experiments with multiple independent variables: a resource management perspective on complete and reduced factorial designs,” *Psychological Methods*, 14, 202–224.
- Dasgupta, T., Pillai, N., and Rubin, D. B. (2015), “Causal inference from  $2^K$  factorial designs by using potential outcomes,” *Journal of the Royal Statistical Society: Series B*, 47, 727–753.
- Ehrenfeld, S. (1955), “On the efficiency of experimental designs,” *The Annals of Mathematical Statistics*, 26, 247–255.

- Fisher, R. A. (1925), *Statistical Methods for Research Workers*, Oliver & Boyd: Edinburgh.
- Gadbury, G. L. (2001), “Randomization inference and bias of standard errors,” *The American Statistician*, 55, 310–313.
- Gupta, S. and Mukerjee, R. (1989), *A Calculus for Factorial Arrangements*, Springer: New York, NY.
- Hansen, M. H., Hurwitz, W. N., and Madow, W. G. (1953), *Sample Survey Methods and Theory*, Wiley: New York, NY.
- Hedayat, A. S. and Sinha, B. K. (1991), *Design and Inference in Finite Population Sampling*, Wiley: New York, NY.
- Holland, P. W. (1986), “Statistics and causal inference,” *Journal of the American Statistical Association*, 81, 945–960.
- Horvitz, D. G. and Thompson, D. J. (1952), “A generalization of sampling without replacement from finite universe,” *Journal of the American Statistical Association*, 47, 663–685.
- Imbens, G. W. and Rubin, D. B. (2015), *Causal Inference for Statistics, Social and Biomedical Sciences: An Introduction*, Cambridge University Press: New York, NY.
- Kempthorne, O. (1955), “The randomization theory of experimental inference,” *Journal of the American Statistical Association*, 50, 946–967.
- Kish, L. (1965), *Survey Sampling*, Wiley: New York, NY.
- Lu, J. (2016), “On randomization-based and regression-based inferences for  $2^K$  factorial designs,” *Statistics and Probability Letters*, 112, 72–78.
- Neyman, J. (1923/1990), “On the application of probability theory to agricultural experiments. Essay on principles (with discussion). Section 9 (translated),” *Statistical Science*, 5, 465–480.
- (1935), “Statistical problems in agricultural experimentation (with discussion),” *Supplement to the Journal of the Royal Statistical Society*, 2, 107–180.
- Rubin, D. B. (1974), “Estimating causal effects of treatments in randomized and nonrandomized studies,” *Journal of Educational Psychology*, 66, 688–701.
- (1975), “Bayesian inference for causality: The importance of randomization,” in *The Proceedings of the Social Statistics section of the American Statistical Association*, pp. 233–239.
- (1977), “Assignment to treatment group on the basis of a covariate,” *Journal of Educational Statistics*, 2, 1–26.

- (1980), “Randomization analysis of experimental data: The Fisher randomization test comment,” *Journal of the American Statistical Association*, 75, 591–593.
  - (1990), “Neyman (1923) and causal inference in experiments and observational studies,” *Statistical Science*, 5, 472–480.
  - (2008), “Statistical inference for causal effects, with emphasis on applications in epidemiology and medical statistics.” *Handbook of Statistics*, 27, 28–63.
- Sabbaghi, A. and Rubin, D. B. (2014), “Comments on the Neyman-Fisher controversy and its consequences,” *Statistical Science*, 29, 267–284.
- Wilk, M. B. (1955), “The randomization analysis of a generalized randomized block design,” *Biometrika*, 42, 70–79.
- Wilk, M. B. and Kempthorne, O. (1956), “Some aspects of the analysis of factorial experiments in a completely randomized design,” *The Annals of Mathematical Statistics*, 27, 950–985.
- (1957), “Non-additivities in a Latin square design,” *Journal of the American Statistical Association*, 52, 218–236.
- Wu, C. F. J. and Hamada, M. S. (2009), *Experiments: Planning, Analysis, and Optimization*, Wiley: New York, NY.
- Zhao, A., Ding, P., and Dasgupta, T. (2016), “Randomization-based causal inference from unbalanced  $2^2$  split-plot designs,” <http://arxiv.org/abs/1602.03915>.